ABSTRACT:

The invention relates to a method of generating a maximum entropy speech model for a speech recognition system.

To improve the statistical properties of the generated speech model there is proposed that:

5   -   by evaluating a training corpus, first probability values $p_{ind}(w \mid h)$ are formed for N-grams with $N \geq 0$;

     -    an estimate of second probability values $p_{\lambda}(w \mid h)$, which represent speech model values of the maximum entropy speech model, is made in dependence on the first probability values;

10   -   boundary values m_{α} are determined according to the equation

$$m_{\alpha} = \sum_{(h,w)} p_{ind}(w \mid h) \cdot N(h) \cdot f_{\alpha}(h,w)$$

where $N(h)$ is the rate of occurrence of the respective history h in the training corpus and $f_{\alpha}(h, w)$ is a filter function which has a value different from zero only for certain N-
15   grams predefined a priori and featured by the index α, and otherwise has the zero value;

     -    an iteration of speech model values of the maximum entropy speech model is continued until values $m_{\alpha}^{(n)}$ determined in the nth iteration step according to the formula

20

$$m_{\alpha}^{(n)} = \sum_{(h,w)} p_{\lambda}^{(n)}(w \mid h) \cdot N(h) \cdot f_{\alpha}(h,w)$$

sufficiently accurately approach the boundary values m_{α} according to a predefinable convergence criterion.